DOCUMENT RESUME

ED 465 073 CE 083 325

AUTHOR	Austin, James T.; Mahlman, Robert A.
TITLE	High-Stakes Testing: Implications for Career and Technical Education. The Highlight Zone: Research @ Work.
INSTITUTION	National Dissemination Center for Career and Technical Education, Columbus, OH.
SPONS AGENCY	Office of Vocational and Adult Education (ED), Washington, DC.
REPORT NO	NDCCTE-7
PUB DATE	2002-00-00
NOTE	10p.; In: Research @ Work Number 6, see ED 461 730.
CONTRACT	V051A990004
AVAILABLE FROM	NDCCTE Product Sales Office, Ohio State University,1900 Kenny Road, Columbus, OH 43210-1090; Tel: 800-678-6011, ext. 24277 (Toll Free); Tel: 614-292-4277; Fax: 614-688-3258; Tel: 614-292-1260; e-mail: ndccte@osu.edu; Web site: http://www.nccte.org. For full text: http://www.nccte.org/publications/infosynthesis/highlightzone/highlight07/highlight07-highstakes.pdf.
PUB TYPE	Information Analyses (070)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Accountability; Adult Education; Educational Research; *High Stakes Tests; Postsecondary Education; Secondary Education; Standardized Tests; State Programs; *Student Evaluation; Test Interpretation; *Test Use; *Test Validity; Testing Programs; *Vocational Education
IDENTIFIERS	*Career and Technical Education

ABSTRACT

The topic of high-stakes testing (HST) is important because HST has direct and indirect effects on career-technical education (CTE) programs and timely because HST increasingly enters public discussion and has produced a large body of research and practice that generalizes to CTE. A review of HST has identified two persisting dilemmas: policy and public expectations of testing exceeding tests' technical capacities and tension between testing to increase fairness and testing to classify. Applicable strategies to provide validity for HST are reliability estimation of scores used to make decisions, expert judgements of item linkage to curricula, studies of the predictive power of HST scores, and studies of consequences. Two opposing perspectives on the accountability-testing theme are that use of HST for accountability is a positive application of data-driven management to education and that the consequences of HST are negative. Descriptions of HST systems in Kentucky, Texas, and Massachusetts indicate different ways to accomplish HST; use of advisory panels to represent stakeholders' viewpoints; and continuous change. Findings of an e-mail survey of state CTE directors suggest ways to expand assessment modalities--computer delivery of assessments and authentic assessments or multimodal assessments that include high- and low-stakes components. Implications are that the CTE community needs awareness of HST; tests should be used responsibly; and a useful database system should be developed. (Contains 76 references and a list of 8 Internet sites.) (YLB)

ED 465 073

High Stakes Testing: Implications for Career and Technical Education

James T. Austin
Robert A. Mahlman

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

CE083325

THE HIGHLIGHT ZONE
Research @ Work
no. 7
CTE National Dissemination Center
1900 Kenny Road, Columbus, OH 43210-1090

High-Stakes Testing: Implications for Career and Technical Education

by James T. Austin and Robert A. Mahlman

2002

High-stakes testing systems generate scores with important consequences, which are applied to students in the form of remedial course work or graduation requirements. In addition, these scores are being used at classroom and district levels to evaluate teacher and administrator performance. Thus, the topic of high-stakes testing (HST) is important and timely: important because HST has direct and indirect effects on career-technical programs at all levels; timely because HST increasingly enters public discussion and has produced a large body of research and practice. There is a need to understand HST given the consequences for students, teachers, and administrators. We located little research on HST that *speaks directly* to career-technical education (CTE). Two points, however, are relevant to the scarcity. First, we believe that the existing research generalizes to CTE populations. Second, the lack supports the need for a researcher-practitioner dialog advocated by Seashore-Louis and Jones (2001). Our goal is to provide a balanced interpretation of research and practice on HST for the CTE community. In order to do that, this document is divided into four sections. The sections focus on (1) establishing the context of testing in education and in career-technical education, (2) reviewing trends and issues of HST, (3) presenting snapshots of HST systems in three states—Texas, Massachusetts, and Kentucky—to illustrate approaches to validity, cost, and fairness; and (4) developing the implications of HST for career-technical education. We turn first to the context of testing.

Context of Testing in Education and CTE

We develop the context through a brief, selective presentation of major reviews of validity. Two comprehensive reports were published by the Office of Technology Assessment (OTA): ***Testing in American Schools*** (1992) and ***Testing and Assessment in Vocational Education*** (1994). The 1992 report is a general look at educational testing. The purposes of testing are identified as classroom feedback, system monitoring, and selection-placement-credentialing. A summary and policy options are presented. Policies presented in the form of budget scenarios are rank ordered on cost. The least expensive approach was keeping educational testing dollars at roughly 7% of $100 million in educational research appropriations. The most costly approach was support for test development research on linkages between testing and cognitive science, consensus-building techniques for test content, generalizability of new testing methods, and validation of new testing methods. Intermediate cost is represented by creating a clearinghouse to synthesize test use research, providing professional development for teachers in new assessment methods, and building a national database of test items. The effects of the recently passed No Child Left Behind Act on such scenarios remain to be seen.

Testing and Assessment in Vocational Education (OTA 1994) was organized around (1) purpose, (2) performance-based accountability in federal law, (3) state policies on testing, (4) studies of vendors of occupational competency tests, (5) broad technical skills, and (6) implementation of performance standards and measures. On the basis of a 1992-1993 survey, the authors placed state CTE assessment policies into four categories: 18 states mandated or strongly encouraged local written competency testing for occupational skills, 15 states mandated assessment of occupational skills in local programs without specifying how, 10 states encouraged assessment of occupational skills in local programs without specifying how, and 8 states had no specific policy about testing. Skill categories for assessment include academic, vocational, generic workplace, and broad technical (communications). The conclusions were fourfold:

1. Assessment practices in secondary CTE differed considerably from the rest of education. The best practices resembled the alternative forms of assessment just then being explored for the rest of education, but the quality of these assessments varied greatly.
2. In CTE, testing is not an after-the-fact, external process of inspection but integral to education—a goal just now being advanced in academic education.
3. Critical issues in performance assessment are the comparability of judgments (across instructors and programs) and correspondence of the judgments with standards.
4. Critical issues in written testing are (a) the relevance of test items to capabilities for job performance and (b) the long-term effects of the testing method on teaching and learning.

Stecher et al. (1997) studied six alternative assessment systems in CTE: (1) Career-Technical Assessment Program, (2) Kentucky Instructional Results Information System, (3) Laborers-AGC Environmental Training and Certification, (4) National Board for Professional Teaching Standards Certification, (5) Oklahoma Competency-Based Testing Program, and (6) Vocational Industrial Clubs of America (VICA) National Competition. A framework for choice of assessment was presented based upon *purposes of assessment* (improve learning, certify individual mastery, evaluate program success), *quality* (validity, fairness) and *feasibility* (cost, credibility). A fourth facet is the *context of CTE*, in which student characteristics and program content drive assessment choice. Stecher and colleagues raised other issues, including number of measures, stakes, type of tasks, standardization, number of purposes, and participation. They reached several conclusions. Primarily, they argued (and we agree) that alternative assessments are useful tools for CTE. They suggested considering the three purposes within the factors of context, quality, and feasibility. Clearly, there is no best assessment that crosses the three purposes and contexts (urban vs. rural or secondary vs. postsecondary). Their review suggests that performance assessments or portfolios can function as stand-alone assessments and as components of assess-

BEST COPY AVAILABLE

ment packages. Relatively few HST systems use alternative assessments.

Heubert and Hauser (1999) published a comprehensive review of HST for the National Research Council. Their focus was tests used to make decisions about individuals, including tracking and placement, promotion and retention, and awarding or withholding high school diplomas. A panel of experts reviewed controversies that may emerge when test scores can open or close gates on educational pathways. The panel organized their work around the following themes: (1) judging appropriateness of tests, (2) making tests reliable, valid, and fair, (3) advancing and promoting proper test use; and (4) recommending how decision makers in education should—and should not—use test results. Two persisting dilemmas were identified. The first is that policy and public expectations of testing often exceed the technical capacities of tests (leading to test use for nonvalidated purposes). The second dilemma is a tension that exists between testing to increase fairness and testing to classify.

The major quality issue for any test score is validity, which refers to support for desired interpretations. This topic and others are treated extensively in the recent revision of the ***Standards for Educational and Psychological Testing*** (American Educational Research Association et al. 1999). The "stakes" of testing are directly related to validity requirements; thus it is incumbent upon developers and users of tests to provide strong evidence for high-stakes tests. Haertel (1999), in a discussion of validity for HST, noted that validation flows from the intended purpose of the assessment (How will a score be interpreted?). He suggested several designs for examining the validity of high-stakes tests. Some of the applicable evidence strategies for HST include reliability estimation of those scores used to make decisions (overall or components), expert judgments of item linkage to curricula, studies of the predictive power of HST scores against further education and labor market criteria, and studies of the consequences (intended and unintended) of HST. Related to consequences, research just published suggests that one consequence might be an increase in dropout rates. This increase might be especially problematic because some studies suggest that it occurs at the lower ability levels (Jacob 2001; Roderick and Engel 2001). Although there are possible negative consequences, Cizek (2001) reviewed 10 unintended consequences of HST systems that are positive. Finally, in a country-level analysis, Bishop (2000) reported that the use of curriculum-based exit exams was associated with increased learning through various individual and school system mechanisms (student reports of effort, district teacher hiring practices).

Trends and Issues in HST

What are some trends and issues in HST? Testing worldwide is increasing (Airasian 1987; Linn 2000; Madaus 1995). HST systems are expanding in most states. The expansion includes testing for career-technical students as well as students with disabling conditions (Langenfeld, Thurlow, and Scott 1996). Phelps (2000), an advocate of testing, provides data on the prevalence of testing across the countries of the Organization for Economic Cooperation and Development and the "demand" on the part of U.S. stakeholders for standardized testing (Phelps 1998). In his summary of surveys, he concludes that a majority of Americans are positive, over time, about testing.

Scores from HST systems now dominate accountability programs (Adams and Kirst 1999; Linn 2000). Major tools of accountability are standards and assessments. Standards can be grouped into content and performance types (Resnick and Wirt 1996). *Content standards* indicate "what" should be learned. They influence curriculum and instruction, they should drive assessment, and they are themselves developed, validated, and revised. *Performance standards* state "how well" the content standards should be learned. Performance standards, within assessment systems, are cutoff scores or benchmarks that form groups of scores associated with levels of mastery. The links between content and performance standards are asserted to drive systemic reform (Marzano and Kendall 1997; Resnick and Wirt 1996; Vinovskis 1996). The American Federation of Teachers (AFT 2001) provides a state-level evaluation of standards, assessments, and their alignment.

There are opposing perspectives on the accountability-testing theme. One is that the use of HST for accountability is a positive application of data-driven management to education. The logic is that, absent information provided by testing to establish baselines and track progress, the enterprise is rudderless. Politicians representing all points of the continuum call for assessment of learners, teachers, and educational systems. Both major party candidates for President during 2000 advocated testing as a means to improve education. The No Child Left Behind Act and the strategic plan of the U.S. Department of Education are clear in their general tone. The act calls for accountability through annual reading and math assessments from grades 3-8. The first two goals of the Department of Education's strategic plan for 2002-2007 are to create a culture of achievement and to improve student achievement. State tests will be benchmarked against the National Assessment of Educational Progress to evaluate quality. Eventually, there will be consequences for schools.

An opposing view is that the consequences of HST are negative. Opposition is found in books against standardized testing, in parental and Internet grassroots organizations, and in media coverage. Significant voices are raised against expansion of standardized testing, a format traditionally valued for its balance of validity and efficiency. One argument is that this format detracts from curriculum and instruction and forces a narrow focus that is devoid of critical thinking. These beliefs are expressed in books critical of testing (Kohn 2000; Popham 2000; Sacks 2000). FairTest (www.fairtest.org) studies and disseminates material on the inadequacies of standardized testing, including the ***Principles and Indicators for Student Assessment Systems*** (1995), which is used to conduct state-level reviews.

Within these perspectives, consider the advantages and disadvantages of HST in Table 1 (Paris 2000). Two caveats are the scarcity of research on the effects of HST and disagreement concerning the support for some of the assertions.

Supplementing a federal focus, a major emphasis on HST flows from the states. A policy paper developed by the Education Commission of the States (2001) reviewed testing and accountability practices state by state. Assessment is clearly a component of systemic educational reform (Vinovskis 1996). Why the popularity? Airasian (1988) used "symbolic validation" to capture the communication inherent in HST. Mandated HST programs, in this view, are supported because they symbolize order and control, desired educational outcomes, and traditional moral values. Along with a steady increase in frequency, the popularity of assessment cycles in American society is noted by Linn (2000). A critique is that the emphasis on HST is due to a "business of testing" complex that is shrinking through mergers. Haney, Madaus, and Lyons

BEST COPY AVAILABLE

Table 1: Advantages and Disadvantages of HST

Selected Advantages

1. Students will work harder and learn more under HST.
2. Students and teachers need HST to know what is important to learn and to teach.
3. HST provides good measurement of the curricula that students are taught in schools.
4. Tests are a "level playing field" and provide an equal opportunity for all to demonstrate knowledge.

Selected Disadvantages

1. Traditional tests encourage low-level thinking.
2. Traditional tests misdirect student motivation.
3. Traditional tests discriminate against members of ethnic minority groups.
4. Traditional tests are often not aligned with curriculum.

Source: Paris (2000)

(1993) suggested that one outcome of concentration in this sector is reduced test quality. Their concern about quality receives support in a series in the ***New York Times*** (Henriques and Steinberg 2001, Steinberg and Henriques 2001). A matrix was constructed that crossed all states with major test publishers (NCS Pearson, Harcourt Educational Measurement, CTB/McGraw-Hill, Riverside). Then, issues of scoring, analysis, and lateness were reviewed for each state-developer combination. This article suggests increasing concentration in the testing industry beyond that identified by Haney et al. (1993). The concentration occurs at a time when the demand for HST is increasing (Clarke, Madaus, Horn, and Ramos 2001).

State-Level Examples: Kentucky, Texas, Massachusetts

We believe that it is instructive to describe HST systems in states, and our choices were Kentucky, Texas, and Massachusetts. This description can assist the reader in understanding the context of HST efforts. Obviously, detailed state-level comparisons are useful. Reports by the Consortium on Policy Research in Education (http://www.cpre.org/Publications/Publications_Accountability.htm) and by the American Federation of Labor (http://www.aft.org/edissues/standards/msm2001), for example, were consulted. These were useful for determining the scope of HST in general across states. Other sources of archival data covering the states were the Council of Chief State School Officers (http://www.ccsso.org), National Center on Education Statistics (http://www.nces.gov), and FairTest (http://www.fairtest.org). We extracted elements from such reports for our descriptions.

Kentucky

Kentucky has used assessment for accountability since 1990, when the Kentucky Educational Reform Act was passed. The current system is the Commonwealth Accountability and Testing System (CATS). CATS includes a norm-referenced battery (Terra Nova, or Comprehensive Test of Basic Skills 5), which is administered in grades 3, 6, and 9. A standards-based battery (Kentucky Core Content Tests) is administered at the other grade levels. Four levels of performance are defined: Distinguished, Proficient, Apprentice, and Novice. Goals are focused on moving most Kentucky high school students to the proficient level by 2014. The AFT (2001) study reported that the standards were "clear and specific" at elementary, middle, and high school levels for Math and Science, at two levels for Social Studies, and at one level for English. Assessments were aligned at all levels across all domains. Consequences in place are "other incentive" for the secondary level. One assessment contractor is Human Resources Research Organization (HumRRO), another is CTB/McGraw-Hill. Contractors, particularly HumRRO, have conducted extensive research on the validity and impact of performance-based assessments. Researchers have visited school districts and conducted interviews with hundreds of teachers. A consistent theme that emerges from this research is that educators were working hard to understand and adapt to performance-based content demands. A slightly different perspective, however, is presented in a book by Whitford and Jones (2000). That book advances some negative consequences of HST systems for teachers and administrators. Further, Kentucky has moved away from performance-only assessments. This shift away from performance assessment is widely duplicated due to problems with costs and uneven standardization of tests and scoring (Mehrens 1992). Career-technical students in Kentucky are included in testing, and the results are used for federal reporting (Perkins III). Details are provided at the CTE website (http://www.kde.state.ky.us/osis/voced). The Occupational Skills Standards Assessment System was scheduled for roll-out in 2000-2001.

Texas

Education and assessment in the state of Texas received considerable attention during the 2000 presidential campaign. A detailed history of Texas testing systems is provided by Cruse and Twing (2000). Currently, the Texas Academic Assessment System (TAAS) is the HST system. This system employs NCS Pearson as contractor; the programmatic development procedures were described by Smisko, Twing, and Denny (2000). According to the AFT (2001) survey, standards were "clear and specific" at three levels for Math, at two levels for Science and English, and not clear or specific at any level for Social Studies. Assessments were aligned at all three levels for English and Math, and at two levels for Science and Social Studies domains. The consequences in place are "promotion policies" for elementary/middle school levels and exit exam plus "other incentive" for the secondary level. Procedures used to develop and maintain the TAAS are available at the website of the Texas Education Agency (www.tea.state.tx.us/student.assessment). Career and Technology Education (CATE) students in Texas were under the TAAS and will be under the next generation. The CATE website does not devote extensive space to the TAAS. There may be internal mechanisms for disaggregating by special needs and "track."

One avenue of challenge to HST systems is legal. A lawsuit was filed in 1997 by the GI Forum (Hispanic and African-American stakeholders) to challenge the TAAS on content-curricular links and opportunity to learn. The case was decided in favor of the state by a federal district court judge (Haney 2000). Following the lawsuit, articles in two journals captured the opposing per-

BEST COPY AVAILABLE

spectives. In *Applied Measurement in Education* (vol. 13, no. 4, 2000), the state's side is presented in overviews by Texas Education Agency staff and in articles on validity evidence (Phillips 2000) and experiences of expert consultants (Mehrens 2000, Schaefer 2000). On the other side, the *Hispanic Journal of Behavioral Science* (vol. 22, no. 4, 2000) presented the plaintiff's side. Neither special issue included the opposing viewpoint. Obviously, HST systems are not immune from challenge (Mehrens and Popham 1992; Office of Civil Rights 1999). Some challenges are undoubtedly the result of failing to communicate clearly to stakeholders, but others reflect deeper divisions. Equity is a major concern (Scheuneman and Oakland 1998). One legal argument is that "opportunity to learn" issues favor higher-socioeconomic-status districts. Further challenges will employ assertions of funding inequities, test validity deficits, and reduced opportunity to learn.

Massachusetts

The Massachusetts Comprehensive Assessment System (MCAS) is the state-mandated HST instrument for high school students. The Consortium on Policy Research in Education (2000) report indicates that the MCAS, first administered in 1998, was implemented in response to the Education Reform Law of 1993. That act required MCAS to be designed to (1) test students educated with public funds across the Commonwealth, including students with disabilities and students with limited English proficiency (LEP); (2) be administered annually in at least grades 4, 8, and 10; (3) measure performance based on the Massachusetts Curriculum Framework learning standards; (4) report performance of students, schools, and districts; and (5) serve as one basis of accountability for students, schools, and districts (for example, beginning in 2003, grade 10 students must pass the grade 10 tests as one condition of eligibility for a high school diploma). The MCAS evaluates student knowledge at grades 4, 8, and 10 in the following subjects: Language Arts, Mathematics, Science/Technology and History/Social Studies (only grades 8 and 10). There is some validation evidence, most notably a 1999 validity study that found correlations between the MCAS and a norm-referenced test. The AFT (2001) survey indicates that standards are "clear and specific" at elementary, middle, and high school levels for English, Math, and Science and at two levels for Social Studies. The assessments are aligned at three grade levels for English, Math, and Science, and aligned at two levels for Social Studies. Two consequences, exit exam and "other incentive," are in place for the secondary level.

One reason that we selected this state is that the Massachusetts CTE community is actively attempting to modify the act regulating HST. The Massachusetts Association of Vocational Administrators took a public stance on expanding the HST system. Their three-part solution is to test CTE students in 11th grade to permit vocational-academic integration to operate, to add assessment formats for different learning styles (asserted to be experiential for CTE students), and to delay penalties until an opportunity-to-learn interval, such as 2 years, has passed. The CTE community also tried to have the trade-based Certificate of Occupational Proficiency (COP) substitute for the academic competency determination provided by the MCAS, a proposal rejected by the Commissioner of Education in April 2001. Some scores of CTE students for several different levels can be obtained from the website operated by the state Department of Education (http://www.doe.mass.edu/mcas).

Summary

What may be concluded from the review of these three states? Primarily, there are different ways to accomplish high-stakes testing and associated choices that differ across these three states. Kentucky includes, in different years, norm-referenced (Terra Nova) and criterion-referenced (Kentucky Core Content) tests in its system. Texas documents its assessments well and releases the complete assessment each year after testing and has been through a contentious legal challenge. The states all employ advisory panels to represent the viewpoints of multiple stakeholders (testing specialists, parents, teachers). The situation in Massachusetts represents the most concerted action by the CTE community to influence the legislation, and it also shows the difficulties of influencing HST systems. Kentucky and Massachusetts are featured as "dilemmas" in the report by the Educational Commission of the States (Doumay 2000). It is difficult to disaggregate CTE students in reporting. Lastly, there is continuous change, for example, the progression from TAAS to the Texas Assessment of Knowledge and Skills in 2001.

Conclusions: Implications for CTE

As noted earlier, we found little material published in CTE sources that speaks directly to HST. One way that we developed information was to survey CTE policy makers, as described next.

E-mail Survey of State Directors

We conducted an e-mail survey of state directors on April 19, 2001, using addresses available from the National Association of State Directors of Vocational-Technical Education. The survey asked each state director to reply to the following two-part question: (1) What percentage of your CTE students participate in HST? (0-100%) and (2) What do you and your staff believe are the two or three major implications of HST for CTE?

The number of replies received was 20, a response rate of 40%. The responses to the two items were as follows. For the Percentage of CTE Students Participating in HST, the responses indicated *all or none* with 0% reported by 6 states and 100% reported by 13 states. One respondent indicated 20-30% of the students participated in HST. The implications of HST were investigated through a content analysis of 38 implications provided by respondents. The implications were categorized as (1) positive (55%), (2) negative (34%), and (3) neutral (11%). Sample positive statements included (1) can validate that CTE students are as capable as other students, (2) can provide credibility and accountability for CTE programs, and (3) can create a greater focus on what students should know and be able to do. On the other hand, sample negative statements included (1) remediation will seriously affect CTE enrollments, (2) can consume more of scarce time and resources, and (3) increases graduation requirements, which forces students out of CTE due to scheduling.

Based on the material reviewed in this document, we share several conclusions. One fundamental point of emphasis is that, as HST becomes institutionalized as part of the educational terrain, the CTE community should be aware of quality control. The issues range from insufficient validity evidence to support an operational purpose of testing to fairness in opportunity to learn from a new curriculum before consequences. Many quality issues stem from a failure to use systems thinking. A systems perspective places the HST assessment into context as one component among many.

BEST COPY AVAILABLE

Other components include content standards, curriculum and instruction derived from the content standards, the benchmarks or cut-scores used to divide examinees into groups with associated consequences, and the policies surrounding testing (retesting, accommodations, disaggregation, remediation). Any weaknesses in other components attenuate the advantages of HST. Problematic relationships among components attenuate expected benefits, for example, alignment of standards, curriculum, and assessments (AFT 2001; Glatthorn 1999; Wraga 1999). One way to address this problem is to develop standards for educational accountability systems. Baker and Linn (1999) presented the issue in "Watching the Watchers." Baker (2000) discussed six descriptors for consideration in such systems, and Linn (2001) gave a comprehensive account of designing and developing assessment-accountability systems.

What are the implications of HST for CTE? State directors provided some insight into this question. Recall that slightly over half of the implications provided were positive, but a substantial minority were negative or neutral. Clearly, there is ambivalence, and this ambivalence matches the findings in the research literature. In many states CTE students are now in the same "high-stakes kettle" as students in other tracks. On the one hand, it is good to include all students in large-scale assessments if the intent is to develop baselines and disaggregated reports. One hypothesis is that CTE is an area to which students with low test scores or special needs are steered. Consider a study reported by Elliot, Knight, Foster, and Franklin (2001). They used multiple regression to analyze 3 years of HST scores for about 2,500 Arizona students in both academic and career-technical tracks. Raw scores indicated that CTE students scored lower as a group. However, when special population designations (handicapped, LEP, economic disadvantage, academic disadvantage, and single parent status), learning style (visual vs. kinesthetic), and demographics were included in the analysis, the difference decreased. Those special population designations were strongly correlated with CTE status. More research along this line is needed.

Are there alternatives to traditional HST for the CTE community? Recall that traditional refers to multiple-choice format, which has been valued most for efficiency (cost), to a moderate degree for its validity, and very little for authenticity (Wiggins 1998). Several avenues of expansion are possible. One involves maintaining a multiple-choice format, but adding novel item types. Haladyna (1999) published a book on developing and validating multiple-choice items in which he advocates items that can get at high-level thinking. These include multiple true-false and context-dependent item sets (scenarios), relatively easy methods of assessing higher-level thinking skills. This approach demands a refocusing of test construction and could be implemented through a change in test specifications. There are, however, additional avenues.

Another way to expand would use computer delivery of assessments, either using a stand-alone system or networked systems (Bennett 2002; Kerka and Wonacott 2000). The capabilities of the computer permit extensive use of graphics, audio, and video clips, as well as dynamic assessments that adapt to the test-taker. The use of technology and rapid scoring may facilitate student motivation; these features certainly can influence curriculum planning. Bunderson, Inouye, and Dillon (1989) identified four generations of computerized testing: computerized testing, adaptive testing, continuous measurement, and intelligent measurement. The first two generations are in widespread use, as shown by Dragow and Olson-Buchanan (1999). The latter two generations, however, have not been widely applied. In fact, they may require significant advances in curriculum and assessment theory even to demonstrate their utility. Nichols and Sugrue (1999), for example, document a missing link between test development and cognitive theory.

However, examples of work that links assessment to cognitive models and to technology are appearing. Pellegrino (2001) presents compelling arguments for redesigning education assessment by merging cognition, technology, and measurement. His aspiration is for assessments that are aligned vertically by levels of the educational system; horizontally across assessment, curriculum, and instruction; and over the interval that individuals spend in the (CTE) system. Gott and Lesgold (2000) reviewed cognitive performance models for a specific domain: complex machine troubleshooting. They showed how cognitive analysis of such domains provides useful products, which include performance models, progression of performance models from initial to mastery, and individual differences. There are obvious parallels between their discussion and the content of career-technical education. Shaw, Effken, and Fajen (1997) developed an unobtrusive online method for studying problem-solving paths. Wilson and Sloane (2000) developed a computer system, Berkeley Evaluation and Assessment Research, that embeds assessment within instructional content. This system promises to make instruction and assessment seamless, and its developers use the latest developments in psychometric theory. The system illustrates continuous measurement, in which assessment is integrated into curriculum.

Such systems potentially point the way to adapting instruction *and* assessment. At present, there are software platforms that can implement the first three generations, pointing out the need to develop intelligent measurement (Bunderson et al. 1989). This generation implies the ultimate, which is tailoring both curriculum and assessment to individuals.

A third strategy for expansion involves authentic assessments or multimodal assessments that include high- and low-stakes components. In discussing broad technical skills, OTA (1994) identified five alternative approaches founded on different assumptions about relationships between general and specific skills and between foundational and advanced skills. The alternatives are vocational aptitudes, core occupational skills, occupational maps, design and technology, and cognitive skills. Roeber (1998), however, notes challenges in using innovative assessments.

The CTE community must become aware of assessment standards and position statements that bear directly on HST. In order to become knowledgeable, several sources are relevant. The ***Standards for Educational and Psychological Testing*** (AERA et al. 1999) are the dominant standards, developed collaboratively by three professional organizations and endorsed internally by each organization. The chapter on educational assessment presents 19 standards, and all but a few are relevant to HST. The AERA position on HST, a set of 11 principles, is most definitive because of its direct focus and its close relationship to the 1999 Standards. Hauser, Martin, Qualls, Neill, and Potter (2000) each provided reactions to the AERA principles. Another position was defined by the National Council of Teachers of Mathematics (NCTM 2000) to complement its earlier guidance on assessment (NCTM 1995), whereas the International Reading Association (1999) has come out against HST.

BEST COPY AVAILABLE

A related topic, introduced earlier, is responsible test use. There are several facets, but one is aimed at policy makers. The issue resonates through the efforts of many advocacy groups on both sides of the HST divide. What about accommodations for disabling conditions or for limited English proficiency? There is guidance in reviews of the Washington Assessment of Student Learning (Johnson, Brown, and Kimball 2001) and in Kentucky (Koretz and Hamilton 2000). Heubert and Hauser (1999) reached several conclusions about HST that are evaluative guideposts:

- Accountability responsibility must be *shared* by stakeholders.
- HST should be used only *after* changes ensure opportunity to learn.
- Consequences of HST need *not* be either-or.
- HST should *never* be the only source of information on important decisions.
- Test users should *not* teach narrowly to the test.
- Accuracy of assessment of students with disability or LEP status is tricky.
- The purpose of proposed Voluntary National Tests is *not* to support HST.

Relevant to promoting responsible test use, they described traditional and novel approaches. Traditional methods include professional standards and legal enforcement, whereas novel methods include deliberative forums, independent oversight groups such as the National Board on Educational Testing and Public Policy (Madaus 1992), and federal regulation. The traditional methods are in wide use, but novel ones are proposals that have received some use but are not widespread. The mission of the National Board, for example, is to monitor testing programs, evaluate the benefits and costs of specific testing policies, and evaluate the extent to which professional standards are met in specific contexts. A five-part research agenda is aligned with that mission. Expansion of responsible test use requires thinking about additional stakeholders in HST. If we think about teachers and other educational personnel, one extension concerns assessment literacy defined as the capability to develop, use, and understand assessments and the resulting scores (Bracey 2000; Stiggins 1995). Two worlds have been identified, classroom assessment and statewide assessment, which feature little interaction but could be aligned (Stiggins and Conklin 1992).

Some states have moved toward flexibility. In Ohio, the Governor's Commission for Student Success (2000) was charged with eight tasks. Their fundamental recommendation was that reform requires content standards that are high and realistic. The state legislature passed an act to align the HST system with most recommendations. A major change was phasing out the current system and replacing it, over 6 years, with the Ohio Graduation Test. There will be a 2-year interval between introduction of content standards and testing. A second change was mandating diagnostic tests at certain points. A third change was the creation of a new Educational Management Information System. The system, based on a data warehouse concept, will provide individual, nonidentifiable data.

The value of a useful database system is shown by contrasting the current Ohio system with that used in the Tennessee Value-Added Assessment System (TVAAS). That database stores results from the Tennessee Comprehensive Assessment Program, which uses Terra Nova (i.e., CTBS/5), and it allows tracking of students across grades and teachers. This tracking in turn allows application of a specialized statistical model to estimate student "gains" attributable to the educational system. Such a database is far superior to the Ohio system. Thus, the Tennessee system is rich with implications for practice and research (Sanders and Horn 1998). As an example, analyses have evaluated the relative effects of class size and teacher quality. The latter factor accounts for greater variance in gain scores and also exerts a ripple effect for several years. The deficit created by several years of poor teaching can create a lifetime of lost opportunities. Although valid criticisms of the TVAAS methodology exist (see Linn 2001), the core principles of the system seem desirable. That is, education should add value to a student's repertoire and this repertoire can in part be attributed to actions on the part of teachers and others within educational systems.

In conclusion, the HST movement is now a fixture. The CTE community should try different strategies of engagement to ensure that state-level policy makers receive input from the field. The quality of HST validity evidence should be scrutinized. The ways to expand assessment modalities are worthy of further attention. The unintended consequences of HST systems must receive attention from policy makers and researchers, aided by the CTE community, as data accumulate.

References

Adams, J. E., Jr., and Kirst, M. W. "New Demands and Concepts for Educational Accountability: Striving for Results in an Era of Excellence." In ***Handbook of Research on Educational Administration***, 2d ed., edited by J. Murphy and K. Seashore-Louis, pp. 463-489. San Francisco, CA: Jossey-Bass, 1999.

Airasian, P. W. "State-Mandated Testing and Educational Reform: Context and Consequences." ***American Journal of Education*** 95, no. 3 (May 1987): 392-412.

Airasian, P. W. "Symbolic Validation: The Case of State-Mandated, High-Stakes Testing." ***Educational Evaluation and Policy Analysis*** 10, no. 4 (Winter 1988): 301-313.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. ***Standards for Educational and Psychological Testing***. Washington, DC: AERA, APA, and NCME, 1999. (ED 436 591)

American Federation of Teachers. ***Making Standards Matter 2001***. Washington, DC: AFT, 2001. http://www.aft.org/edissues/standards/MSM2001/Index.htm

Baker, E. L. "Standards for Educational Accountability: SEA Change." Paper presented at the Center for Research on Evaluation, Standards, and Student Testing annual conference, Los Angeles, CA, September 2000.

Baker, E., and Linn, R. "Watching the Watchers: Standards for Accountability Systems." Paper presented at the CRESST annual conference, Los Angeles, CA, September 1999.

Bennett, R. E. ***Using Electronic Assessment to Measure Student Performance. Issue Brief.*** Washington, DC: NGA Center for Best Practices, National Governors' Association, January 2002. http://www.nga.org/cda/files/ELECTRONICASSESSMENT.pdf

Bishop, J. H. "Curriculum-based External Exit Exam Systems: Do Students Learn More? How?" ***Psychology, Public Policy, and Law*** 6 (2000): 199-215.

Bracey, G. W. ***Thinking about Tests and Testing: A Short Primer in "Assessment Literacy."*** Washington, DC: American Youth Policy Forum, 2000. (ED 445 096) http://www.aypf.org/publicatons/BraceyRep.pdf

Bunderson, V.; Inouye, D.; and Olsen, J. B. "The Four Generations of Computerized Educational Measurement." In ***Educational Measurement,*** 3d ed., edited by R. L. Linn, pp. 367-407. New York: Macmillan/ACE, 1989.

BEST COPY AVAILABLE

Cizek, G. J. "More Unintended Consequences of High-Stakes Testing." ***Educational Measurement: Issues and Practice*** 20, no. 4 (Winter 2001): 19-27.

Clarke, M.; Madaus, G.; Horn, C.; and Ramos, M. ***The Marketplace for Educational Testing.*** Boston, MA: National Board on Educational Testing and Public Policy, 2001. (ED 456 146)

Consortium for Policy Research in Education. ***Massachusetts Assessment and Accountability Profile.*** CPRE, 2000. http://www.cpre.org/Publications/ma.pdf

Cruse, K. L., and Twing, J. S. "The History of Statewide Achievement Testing in Texas." ***Applied Measurement in Education*** 13, no. 4 (2000): 327-331.

Dounay, J. ***High-Stakes Testing Systems. ECS StateNotes.*** Denver, CO: Education Commission of the States, March 2000. http://www.ecs.org/clearinghouse/14/56/1456.htm

Drasgow, F., and Olson-Buchanan, J. B., eds. ***Innovations in Computerized Assessment.*** Mahwah, NJ: Erlbaum, 1999.

Education Commission of the States. ***A Closer Look: State Policy Trends in Three Key Areas of the Bush Education Plan—Testing, Accountability and School Choice. Special Report.*** Denver, CO: ECS, 2001. (ED 455 611)

Elliot, J.; Knight, J.; Foster, B.; and Franklin, E. "Are High Stakes Tests a Fair Assessment for CTE Students?" Paper presented at the Association for Career and Technical Education Convention, New Orleans, LA, December 2001.

Glatthorn, A. A. "Curriculum Alignment Revisited: Response to W. G. Wraga." ***Journal of Curriculum and Supervision*** 15, no. 1 (Fall 1999): 26-34.

Governor's Commission for Student Success. ***Expecting More. Higher Achievement for Ohio's Students and Schools.*** Columbus, OH: GCSS, 2000. http://www.osn.state.oh.us/gcss/report.pdf

Gott, S. P., and Lesgold, A. M. "Competence in the Workplace: How Cognitive Performance Models and Situated Instruction Can Accelerate Skill Acquisition." In ***Advances in Instructional Psychology,*** vol. 5, edited by R. Glaser, pp. 239-327. Mahwah, NJ: Erlbaum, 2000.

Haertel, E. H. "Validity Arguments for High-Stakes Testing: In Search of the Evidence." ***Educational Measurement: Issues and Practice*** 18, no. 4 (Winter 1999): 5-9.

Haladyna, T. M. ***Developing and Validating Multiple Choice Items,*** 2d ed. Mahwah, NJ: Erlbaum, 1999.

Haney, W. "The Myth of the Texas Miracle in Education." ***Education Policy Analysis Archives*** 8, no. 41 (August 19, 2000). http://epaa.asu.edu/epaa/v8n41/

Haney, W.; Madaus, G.; and Lyons, G. ***The Fractured Marketplace for Standardized Testing.*** Boston, MA: Kluwer, 1993.

Hauser, R. M.; Martin, W.; Neill, M.; Qualls, A. L.; and Portet, A. "Initial Responses to AERA's Position Statement Concerning High-Stakes Testing." ***Educational Researcher*** 29, no. 8 (November 2000): 27-29.

Henriques, D. B., and Steinberg, J. "Right Answer, Wrong Score: Test Flaws Take Toll." ***New York Times,*** May 20, 2001, section 1, p. 1.

Heubert, J. P., and Hauser, W., eds. ***High Stakes: Testing for Tracking, Graduation, and Promotion.*** Washington, DC: National Academy Press, 1999. (ED 439 151)

International Reading Association. ***High-Stakes Assessments in Reading.*** Newark, DE: IRA, 1999. (ED 435 084)

Jacob, B. A. "Getting Tough?: The Impact of High School Graduation Exams." ***Educational Evaluation and Policy Analysis*** 23 (2001): 99-121.

Johnson, E. S., Brown, S. O.; and Kimball, K. "A Statewide Review of the Use of Accommodations in Large-Scale, High-Stakes Assessments." ***Exceptional Children*** 67, no. 2 (Winter 2001): 251-264.

Kerka, S., and Wonacott, M. E. ***Assessing Learners Online. Practitioner File.*** Columbus: ERIC Clearinghouse on Adult, Career, and Vocational Education, the Ohio State University, 2000. (ED 448 285) http://ericacve.org/docs/pfile03.htm

Kohn, A. ***The Case against Standardized Testing.*** Greenwich, CT: Heinemann, 2000.

Koretz, D., and Hamilton, L. "Assessment of Students with Disabilities in Kentucky: Inclusion, Student Performance, and Validity." ***Educational Evaluation and Policy Analysis*** 22, no. 3 (Fall 2000): 255-272.

Langenfeld, K. L.; Thurlow, M. L.; and Scott, D. L. ***High Stakes Testing for Students: Unanswered Questions and Implications for Students with Disabilities. Synthesis Report No. 26.*** Minneapolis: National Center on Educational Outcomes, University of Minnesota, 1996. (ED 415 627)

Linn, R. L. "Assessments and Accountability." ***Educational Researcher*** 29, no. 2 (2000): 4-14.

Linn, R. L. ***The Design and Evaluation of Assessment and Accountability Systems. CSE Technical Report 539.*** Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, University of California, 2001. (ED 455 286)

Madaus, G. F. "An Independent Auditing Mechanism for Testing." ***Educational Measurement: Issues and Practice*** 11, no. 1 (1992): 26-29, 31.

Madaus, G. F. "A Technological and Historical Consideration of Equity Issues Associated with Proposals to Change Our Nation's Testing Policy." In ***Equity and Excellence in Educational Testing and Assessment,*** edited by M. T. Nettles and A. L. Nettles. Boston: Kluwer, 1995.

Marzano, R. J., and Kendall, J. S. ***The Fall and Rise of Standards-Based Education.*** Aurora, CO: Midcontinent Regional Educational Laboratory, 1997. (ED 398 643)

Mehrens, W. A. "Using Performance Assessment for Accountability Purposes." ***Educational Measurement: Issues and Practice*** 11, no. 1 (1992): 3-9, 20.

Mehrens, W. A. "Defending a State Graduation Test: 'GI Forum v. Texas Education Agency.' Measurement Perspectives from an External Evaluator." ***Applied Measurement in Education*** 13, no. 4 (2000): 387-401.

Mehrens, W. A., and Popham, W. J. "How to Evaluate the Legal Defensibility of High-Stakes Tests." ***Applied Measurement in Education*** 5, no. 3 (1992): 265-283.

National Council of Teachers of Mathematics. ***Assessment Standards for School Mathematics.*** Reston, VA: NCTM, 1995. (ED 388 516)

National Council of Teachers of Mathematics. ***High-Stakes Testing. Position Statement.*** Reston, VA: NCTM, 2000. http://www.nctm.org/about/position_statements/highstakes.htm

Nichols, P., and Sugrue, B. "The Lack of Fidelity between Cognitively Complex Constructs and Conventional Test Development Practice." ***Educational Measurement: Issues and Practice*** 18, no. 2 (November 1999): 18-29.

Office of Civil Rights. ***The Use of Tests When Making High-Stakes Decisions for Students: A Resource Guide for Educators and Professionals.*** Washington, DC: Office of Civil Rights, U.S. Department of Education, 1999.

Office of Technology Assessment. ***Testing in American Schools: Asking the Right Questions.*** Washington, DC: OTA, U.S. Congress, 1992. (ED 340 770)

Office of Technology Assessment. ***Testing and Assessment in Vocational Education.*** Washington, DC: OTA, U.S. Congress, 1994. (ED 368 908)

Paris, S. G. "Trojan Horse in the Schoolyard: The Hidden Threats in High-Stakes Testing." ***Issues in Education*** 6, nos. 1-2 (2000): 1-16.

BEST COPY AVAILABLE

Pellegrino, J. W. ***Rethinking and Redesigning Education Assessment***. Denver, CO: Education Commission of the States, 2001. (ED 456 136) http://www.ecs.org/clearinghouse/24/88/2488.htm

Phelps, R. P. "The Demand for Standardized Student Testing." ***Educational Measurement: Issues and Practice*** 17, no. 3 (Fall 1998): 5-23.

Phelps, R. P. "Trends in Large-Scale Testing Outside the United States." ***Educational Measurement: Issues and Practice*** 19, no. 1 (Spring 2000): 11-21.

Phillips, S. E. "'GI Forum v. Texas Education Agency': Psychometric Evidence." ***Applied Measurement in Education*** 13, no. 4 (2000): 343-385.

Popham, W. J. ***Testing, Testing!*** Boston: Allyn and Bacon, 2000.

Principles and Indicators for Student Assessment Systems Cambridge, MA: National Center for Fair and Open Testing (FairTest), 1995. (ED 400 334)

Resnick, L. B., and Wirt, J. G., eds. ***Linking School and Work: Roles for Standards and Assessments.*** San Francisco, CA: Jossey-Bass, 1996.

Roderick, M., and Engel, M. "The Grasshopper and the Ant: Motivational Responses of Low-Achieving Students to High-Stakes Testing." ***Educational Evaluation and Policy Analysis*** 23, no. 3 (Fall 2001): 197-227.

Roeber, E. D. "The Technical and Practical Challenges in Developing Innovative Assessment Approaches for Use in Statewide Assessment Programs." ***Contemporary Education*** 69, no. 1 (Fall 1997): 6-10.

Sacks, P. ***Standardized Minds: The High Price of America's Testing Culture and What We Can Do to Change It.*** Cambridge, MA: Perseus Books, 2000.

Sanders, W. H., and Horn, S. P. "Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research." ***Journal of Personnel Evaluation in Education*** 12, no. 3 (September 1998): 247-256.

Schafer, W. D. "'GI Forum v. Texas Education Agency': Observations for States." ***Applied Measurement in Education*** 13, no. 4 (2000): 411-418.

Scheuneman, J. D., and Oakland, T. "High-Stakes Testing in Education." In ***Test Interpretation and Diversity: Achieving Equity in Assessment***, edited by J. H. Sandoval et al., pp. 77-103. Washington, DC: American Psychological Association, 1998.

Seashore-Louis, K., and Jones, L. M. ***Dissemination with Impact: What Research Suggests for Practice in Career and Technical Education.*** Columbus: National Dissemination Center for Career and Technical Education, the Ohio State University, 2001. https://www.nccte.org/publications/infosynthesis/r&dreport/DissemnationALL_Seashore.pdf

Shaw, R. E., Effken, J. A., and Fajen, B. R. "An Ecological Approach to the On-line Assessment of Problem-Solving Paths: Principles and Applications." ***Instructional Science*** 25, no. 2 (March 1997): 151-166.

Smisko, A., Twing, J. S., and Denny, P. "The Texas Model for Content and Curricular Validity." ***Applied Measurement in Education*** 13, no. 4 (2000): 333-342.

Stecher, B. M., Rahn, M. L., Ruby, A., Alt, N. M., Robyn, A., and Ward, B. ***Using Alternative Assessments in Vocational Education.*** Santa Monica, CA: RAND, 1997. (ED 400 465)

Steinberg, J., and Henriques, D. B. "When a Test Fails the Schools, Careers and Reputations Suffer." ***New York Times***, May 21, 2001, sec. A, p. 1.

Stiggins, R. J. "Assessment Literacy for the 21st Century." ***Phi Delta Kappan*** 77, no. 3 (November 1995): 238-245.

Stiggins, R. J., and Conklin, N. F. ***In Teachers' Hands: Investigating the Practices of Classroom Assessment.*** Albany, NY: SUNY Press, 1992.

Vinovskis, M. A. "An Analysis of the Concept and Uses of Systemic Educational Reform." ***American Educational Research Journal*** 33, no. 1 (Spring 1996): 53-85.

Whitford, B. L., and Jones, K., eds. ***Accountability, Assessment, and Teacher Commitment: Lessons from Kentucky's Reform Efforts.*** Albany, NY: SUNY Press, 2000.

Wiggins, G. ***Educative Assessment***. San Francisco, CA: Jossey-Bass, 1998.

Wilson, M., and Sloane, K. "From Principles to Practice: An Embedded Assessment System." ***Applied Measurement in Education*** 13, no. 2 (2000): 181-208.

Wraga, W. G. "The Educational and Political Implications of Curriculum Alignment and Standards-Based Reform." ***Journal of Curriculum and Supervision*** 15, no. 1 (Fall 1999): 4-25.

Selected Internet Sites of Interest

Center on Education Policy (http://www.ctredpol.org)

Center for Educational Reform (http://www.edexcellence.net)

Center for Fair and Open Testing (FairTest) (http://www.fairtest.org)

Civil Rights Project (http://www.law.harvard.edu/civilrights/)

Consortium on Policy Research in Education (http://www.cpre.org)

Council of Chief State School Officers (http://www.ccsso.org)

Education Commission of the States (http://www.ecs.org)

National Board on Educational Testing and Public Policy (http://www.nbetpp.bc.edu)

The Highlight Zone: Research @ Work is designed to highlight research findings and provide a synthesis of other information sources. The intention is to help practitioners apply and adapt research results for local use.

James T. Austin is Research Specialist 2 and Robert A. Mahlman is Senior Research Specialist and Director of Assessment Services, Center on Education and Training for Employment, the Ohio State University. The following people are acknowledged for their critical review of the manuscript: Stanley Rabinowitz, Co-Director, Assessment and Standards Development Services, WestEd; Robert Schaeffer, Public Education Director, National Center for Fair and Open Testing; and Raúl Soto, Assistant Director, Career-Technical and Adult Education, Ohio Department of Education.

The work reported herein was supported under the National Dissemination Center for Career and Technical Education, PR/Award (No. VO51A990004) as administered by the Office of Vocational and Adult Education, U.S. Department of Education. However, the contents do not necessarily represent the positions or policies of the Office of Vocational and Adult Education or the U.S. Department of Education, and you should not assume endorsement by the Federal Government.

BEST COPY AVAILABLE